Nuffield CESS

Experimental Methods 2016.
Lecture 2: Hypothesis testing, uncertainty, block random assignment

Centre for Experimental Social Sciences

Lecture 2

- Hypothesis testing
- Confidence bounds
- Block random assignment

Observed Outcomes Local Budget

	BUDGET SHARE IF VILLAGE HEAD IS MALE	BUDGET SHARE IF VILLAGE HEAD IS FEMALE
VILLAGE 1	?	15
VILLAGE 2	15	?
VILLAGE 3	20	?
VILLAGE 4	20	?
VILLAGE 5	10	?
VILLAGE 6	15	?
VILLAGE 7	?	30

Centre for Experimental Social Sciences

Potential Outcomes Local Budget

	BUDGET SHARE IF VILLAGE HEAD IS MALE	BUDGET SHARE IF VILLAGE HEAD IS FEMALE	TREATMENT EFFECT
VILLAGE 1	10	15	5
VILLAGE 2	15	15	0
VILLAGE 3	20	30	10
VILLAGE 4	20	15	-5
VILLAGE 5	10	20	10
VILLAGE 6	15	15	10
VILLAGE 7	15	30	15
AVERAGE	15	30	5

TABLE 3.1

Sampling distribution of estimated ATEs generated when two of the seven villages listed in Table 2.1 are assigned to treatment

	Estimated ATE	Frequency with which an estimate occurs
	−1	2
	0	2
	0.5	1
	1	2
	1.5	2
	2.5	1
	6.5	1
	7.5	3
	8.5	3
	9	1
	9.5	1
	10	1
	16	1
Average	**5**	
Total		**21**

Example: Based on the numbers in Table 3.1, we calculate the standard error as follows:

$$Sum\ of\ squared\ deviations$$

$$= (-1-5)^2 + (-1-5)^2 + (0-5)^2 + (0-5)^2 + (0.5-5)^2 + (1-5)^2 + (1-5)^2$$

$$+ (1.5-5)^2 + (1.5-5)^2 + (2.5-5)^2 + (6.5-5)^2 + (7.5-5)^2 + (7.5-5)^2$$

$$+ (7.5-5)^2 + (8.5-5)^2 + (8.5-5)^2 + (8.5-5)^2 + (9-5)^2 + (9.5-5)^2$$

$$+ (10-5)^2 + (16-5)^2 = 445$$

$$Square\ root\ of\ the\ average\ squared\ deviation = \sqrt{\frac{1}{21}(445)} = 4.60$$

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\left\{\frac{m\mathrm{Var}(Y_i(0))}{N-m} + \frac{(N-m)\mathrm{Var}(Y_i(1))}{m} + 2\mathrm{Cov}(Y_i(0), Y_i(1))\right\}}$$

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{6}\left\{\frac{(2)(14.29)}{5} + \frac{(5)(42.86)}{2} + (2)(7.14)\right\}} = 4.60.$$

$$\widehat{SE} = \sqrt{\frac{\widehat{Var}(Y_i(0))}{N-m} + \frac{\widehat{Var}(Y_i(1))}{m}},$$

Hypothesis Testing

- We can test certain conjectures that provide us a complete schedule of potential outcomes
- One such conjecture is the *sharp null hypothesis* that the treatment effect is zero for all observations
- Under this hypothesis, $Y_i(1) = Y_i(0)$, in which case we observe *both* potential outcomes for every observation
- Simulated randomizations provide an exact sampling distribution of the estimated average treatment effect under the sharp null hypothesis

Centre for Experimental Social Sciences

Observed Outcomes Local Budget

	Budget share if village head is male	Budget share if village head is female
Village 1	?	15
Village 2	15	?
Village 3	20	?
Village 4	20	?
Village 5	10	?
Village 6	15	?
Village 7	?	30

Example: Randomization

- From this table generate an estimate of the ATE of 6.5
- How likely are we to obtain estimate as large as or larger than 6.5 if the true effect were zero for all observations?
- The probability, or p-value, of interest in this case addresses a *one-tailed hypothesis*, namely that female village council heads increase budget allocations to water sanitation
- All applied analyses will be conducted in R

Centre for Experimental Social Sciences

Example: Randomization

- Based on the observed outcomes in Table, we may calculate the 21 possible estimates of the ATE that could have been generated if the null hypothesis were true: {-7.5, -7.5, -7.5, -4.0, -4.0, -4.0, -4.0, -4.0, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, 3.0, 3.0, 6.5, 6.5, 6.5, 10.0, 10.0}.

- How likely are we to obtain an estimate as large as or larger than 6.5 if the true effect were zero for all observations?

Example: 1-tailed test

- The probability, or p-value, of interest in this case addresses a *one-tailed hypothesis*, namely that female village council heads increase budget allocations to water sanitation
- Five of the estimates are as large as 6.5. Therefore, when evaluating the one-tailed hypothesis that female village heads *increase* water sanitation budgets, we would conclude that the probability of obtaining an estimate as large as 6.5 if the null hypothesis were true is 5/21 = 24%.

Example: 2-tailed test

- If we sought to evaluate the *two-tailed hypothesis* – whether female village council heads either increase or decrease the budget allocation for water sanitation

- We would calculate the p-value of obtaining a number that is greater than or equal to 6.5 or less than or equal to -6.5. A two-tailed hypothesis test would count all instances in which the estimates are at least as great as 6.5 *in absolute value*. Eight of the estimates qualify, so the two-tailed p-value is 8/21=38%.

Centre for Experimental Social Sciences

```r
rm(list=ls(all=T))

## ------------------------------------------------
## BASICS: LOOPS AND ITERATIVE PROCESSING
## ------------------------------------------------

## A simple loop
## This code says:
##1. Create a 'counter' variable i
##  2. Initialize the variable so its value is 1
##  3. Conduct the procedures written in the loop
##   and whenever you come across i , replace
##   it with its value, 1
##  4. At the end of the first iteration, replace
##   the value of i with the next value in the series
##   which is 2
##5. Repeat - conduct procedures in the loop
##   treating i as 2, and so on until we reach
##   the max value specified (in this case 5)

for(i in 1:5){

 print("This is the start of iteration number: ")
 print(i)

}
```

```r
## You can iteratively conduct a set of procedures over
## objects defined outside the loop (and before the loop)
## by treating `i' as the value of the index position
## that iteratively updates.

## Here is an example using vectors


odd <- c(1,3,5,7,9)         # given vector
even <- rep(NA,length(odd))   # create shell object for output

for(i in 1:length(odd)){

 even[i] <- odd[i]+1

}

cbind(odd, even) # print old (odd) vs new (even) vectors side-by-side


## You can nest loops -- Here we loop over rows and columns

mat <- matrix(1:12, nrow=3, ncol=4)
mat

for(i in 1:nrow(mat)){
 for(j in 1:ncol(mat)){

   print(paste("Let's print the number in row",i,"and column",j,sep=" "))
   print(mat[i,j])

 }
}
```

Comments on Randomization

- One obtains arbitrarily precise p-values without relying on distributional assumptions

- The same method can be used for a wide variety of applications and test statistics (e.g., the difference-in-means estimator, regression, difference-in-variance, etc.)

- It forces the researcher to take a moment to think carefully about what the null hypothesis is and how it should be tested.

Confidence Intervals

- We cannot estimate the dispersion of the estimates without making simplifying assumptions
- A simple approach is to assume that the treatment effect for every subject is equal to the estimated ATE
- For subjects in the control condition, missing $Y_i(1)$ values are imputed by adding the estimated ATE to the observed values of $Y_i(0)$.
- For subjects in the treatment condition, missing $Y_i(0)$ values are imputed by subtracting the estimated ATE from the observed values of $Y_i(1)$
- This approach yields a complete schedule of potential outcomes, which we may then use to simulate all possible random allocations.

Effect of winning visa lottery on attitudes toward people from other countries

- We cannot estimate the dispersion of the estimates without making simplifying assumptions
- Winners and losers were asked to rate the Saudi, Indonesian, Turkish, African, European, and Chinese people on a five-point scale ranging from very negative (-2) to very positive (+2).
- Adding the responses to all six items creates an index ranging from -12 to +12
- Average in the treatment group is 2.34
- 1.87 in the control group

Centre for Experimental Social Sciences

Pakistani Muslims Lottery

RATINGS OF PEOPLE FROM OTHER COUNTRIES	CONTROL (%)	TREATMENT (%)
-12	0	0.2
-9	0.22	0
-8	0	0.2
-6	0.45	0.2
-5	0	0.2
-4	0.45	0.59
-3	0	0.2
-2	1.12	0.98
-1	1.56	2.75
0	27.23	18.63
1	18.3	13.14
2	24.33	25.29
3	8.48	10.98
4	5.8	9.61
5	3.35	3.92
6	3.79	7.25
7	2.23	2.55
8	0.89	1.37
9	0.22	0.78
10	0.45	0
11	0.67	0.2
12	0.45	0.98
TOTAL	100	100
N	(448)	(510)

Centre for Experimental Social Sciences

Estimate our 95% interval

- We add 0.47 to the observed outcomes in the control group in order to approximate their values;
- We subtract 0.47 from the treatment group's observed outcomes in order to approximate their values
- Simulating 100,000 random allocations using this schedule of potential outcomes and sorting the estimated ATEs in ascending order
- We find that the 2,500th estimate is 0.16 and the 97,501st estimate is 0.79, so the 95% interval is [0.16, 0.79]

Centre for Experimental Social Sciences

```
##########################################
 ##
# R Starter Code - Randomization Inference
# POLS 4368 Section - Feb 12 2013
##########################################
 ##

####################################
## Load the RI package
####################################

# install.packages("ri",
 dependencies=TRUE)
library(ri)
set.seed(1234567)

####################################
## Generate data, or read-in data
####################################
N <- 50
m <- 25

d <- ifelse(1:N %in% sample(1:N, m), 1, 0)
Y0 <- runif(N,0,1)
Y1 <- Y0 + rnorm(N,2,2)
Y <- Y1*d + Y0*(1-d)

cbind(Y0,Y1,d,Y)  # look at your data

## Conduct analysis of actual experiment
## Estimate the ATE

# nonparametric
mean(Y[d==1]) - mean(Y[d==0])

# or fitting data to ols
lm(Y~d)
```

```r
# Define inputs (Z, Y, any blocking variable, or pre-treatment variables)
# Z must be a binary variable 0=control, 1=treatment
Z <- d


probs <- genprobexact(Z)
ate <- estate(Y,Z,prob=probs)

# Set the number of simulated random assignments
perms <- genperms(Z,maxiter=10000)

# Create potential outcomes UNDER THE SHARP NULL OF NO EFFECT FOR ANY UNIT
Ys <- genouts(Y,Z,ate=0)

# Generate the sampling distribution based on schedule of potential outcome
# implied by the sharp null hypothesis
distout <- gendist(Ys,perms,prob=probs)

ate                                    # estimated ATE
sum(distout >= ate)                    # one-tailed comparison used to calculate p-value (greater than)
sum(abs(distout) >= abs(ate))          # two-tailed comparison used to calculate p-value

dispdist(distout,ate)                  # display p-values, 95% confidence interval, standard error under the null, and graph the sampling distribution under the null

#---------------------------------------------------------------
# estimation of confidence intervals assuming ATE=estimated ATE
#---------------------------------------------------------------
Ys <- genouts(Y,Z,ate=ate)               # create potential outcomes UNDER THE ASSUMPTION THAT ATE=ESTIMATED ATE

distout <- gendist(Ys,perms,prob=probs)  # generate the sampling distribution  based on the schedule of potential outcomes implied by the null hypothesis

dispdist(distout,ate)                  # display p-values, 95% confidence interval, standard error under the null, and graph the sampling distribution under the null
```

Block Random Assignment

- Subjects are partitioned into blocks
- Complete random assignment within each block
- Example
  - Split the sample by gender
  - Select 5 subjects from the male group for the treatment
  - Select 5 subjects from the female group for the treatment

Bertrand and Mullanathan (2004)

	Panel A: Subjective Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (*p*-value)
White names	8.50	10.79	1.27	2.29
	[1,212]	[1,223]		(0.0557)
African-American names	6.19	6.70	1.08	0.51
	[1,212]	[1,223]		(0.6084)

	Panel B: Predicted Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (*p*- value)
White names	7.18	13.60	1.89	6.42
	[822]	[816]		(0.0000)
African-American names	5.37	8.60	1.60	3.23
	[819]	[814]		(0.0104)

Centre for Experimental Social Sciences

Why Block Random Assignment: Practical Concerns

- Program requirements may restrict the number of subjects allowed to receive the treatment
- E.g. summer reading program concerned about students with low levels of preparedness 60% of the admitted students must pass basic skills test
- If 50 students are admitted, randomly select 20 from the applicants that failed and 30 from those who passed
- Fairness concerns require each treatment of demographic groups
- Resource constraints mean you are only able to sample a certain number of subjects from certain regions

Why Block Random Assignment: Statistical Concerns

- Reduces sampling variability
- Subjects in blocks likely to have similar potential outcomes (those who fail and those who pass)
- Especially effective in small samples
- Ensures the ability to do subgroup analysis
    - e.g. men and women
- Complete random assignment may lead to imbalance

Potential Outcomes

VILLAGE	BLOCK	Y(I) CONTROL	Y(I) TREATMENT
1	A	0	0
2	A	1	0
3	A	2	1
4	A	4	2
5	A	4	0
6	A	6	0
7	A	6	2
8	A	9	3
9	B	14	12
10	B	15	9
11	B	16	8
12	B	16	15
13	B	17	5
14	B	18	17

TABLE 3.3

Schedule of potential outcomes for public works projects when audited (Y(1)) and not audited (Y(0))

		All subjects		Block A subjects		Block B subjects	
Village	Block	Y(0)	Y(1)	Y(0)	Y(1)	Y(0)	Y(1)
1	A	0	0	0	0		
2	A	1	0	1	0		
3	A	2	1	2	1		
4	A	4	2	4	2		
5	A	4	0	4	0		
6	A	6	0	6	0		
7	A	6	2	6	2		
8	A	9	3	9	3		
9	B	14	12			14	12
10	B	15	9			15	9
11	B	16	8			16	8
12	B	16	15			16	15
13	B	17	5			17	5
14	B	18	17			18	17
	Mean	9.14	5.29	4.00	1.00	16.0	11.0
	Variance	40.41	32.49	7.75	1.25	1.67	17.0
	Cov(Y(0), Y(1))	31.03		2.13		1.00	

	All Subjects		Block A		Block B	
	Y_i^c	Y_i^t	Y_i^c	Y_i^t	Y_i^c	Y_i^t
Mean	9.14	5.29	4.00	1.00	16.00	11.00
Variance	40.41	32.49	7.75	1.25	1.67	17.00
Covariance	31.03		2.13		1.00	

Estimating the ATE with Block Random Assignment

$$ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$$

- Where J is the number of blocks and N(j)/N is the share of all subjects in block j
- Weighted average of the block-specific ATEs

Observed Outcomes

VILLAGE	BLOCK	Y(I) CONTROL	Y(I) TREATMENT
1	A	0	?
2	A	1	?
3	A	?	1
4	A	4	?
5	A	4	?
6	A	6	?
7	A	6	?
8	A	?	3
9	B	14	?
10	B	?	9
11	B	16	?
12	B	16	?
13	B	17	?
14	B	?	17

Precision of Complete vs. Block Random Assignment

$SE(\widehat{ATE})$ with complete random assignment

$$= \sqrt{\frac{1}{N-1}\left\{\frac{mVar(Y_i^c)}{N-m} + \frac{(N-m)Var(Y_i^t)}{m} + 2Cov(Y_i^c, Y_i^t)\right\}}$$

$$= \sqrt{\frac{1}{13}\left\{\frac{(4)(40.41)}{10} + \frac{(10)(32.49)}{4} + (2)(31.03)\right\}}$$

$$= 3.50$$

$SE(\widehat{ATE})$ with block random assignment

$$= \sqrt{SE_1^2\left(\frac{N_1}{N}\right)^2 + SE_2^2\left(\frac{N_2}{N}\right)^2}$$

$$= \sqrt{(1.23)^2\left(\frac{8}{14}\right)^2 + (2.71)^2\left(\frac{6}{14}\right)^2}$$

$$= 1.36$$

Estimating the ATE with Block Random Assignment

$$
\begin{aligned}
\widehat{ATE} &= (\widehat{ATE_1})\left(\frac{N_1}{N}\right) + (\widehat{ATE_2})\left(\frac{N_2}{N}\right) \\
&= (-1.5)\left(\frac{8}{14}\right) + (-2.75)\left(\frac{6}{14}\right) \\
&= -2.04
\end{aligned}
$$

Standard Error of the Estimated ATE

$$\widehat{SE}(\widehat{ATE}) = \sqrt{\widehat{SE}_1^2 \left(\frac{N_1}{N}\right)^2 + \widehat{SE}_2^2 \left(\frac{N_2}{N}\right)^2}$$

where for each of the two blocks:

$$\widehat{SE} = \sqrt{\frac{\widehat{Var}(Y_i^c)}{N-m} + \frac{\widehat{Var}(Y_i^t)}{m}}$$